ECONOMETRICS EXTRAORDINARY FINAL EXAM

UNIVERSIDAD CARLOS III DE MADRID

JULY 1, 2022

NAME:
NIA:
GROUP:

<u>INSTRUCTIONS</u>:

1. Write your name and group clearly in all sheets.
2. Leave an ID card with your picture on the desk.
3. Each of the four questions will be answered on (both sides of) the sheet where it is written. You cannot use the space on other sheets, or additional sheets.
4. You can use both sides of this sheet only for calculations which will not be evaluated.
5. All parts in each question have the same value.
6. The exam lasts 120 minutes:
   (a) Questions 1, 2 and 3 will be answered in the first 80 minutes using only a pen or pencil.
   (b) Then you can take your personal computer, where the Wooldridge database will have been downloaded in advance. 5 minutes will be given to boot the computer. Question 4 will be answered in 35 minutes using GRETL. Only GRETL can be visible on the computer screen, no other programs can be running. A personal calculator can be used if GRETL's one is not working. Critical values and p-values can be obtained in GRETL.

NAME: ______________________________ GROUP: __________

QUESTION 1 (20%): We wish to estimate an Engel curve for food, relating the proportion of the household income spent in food (Y) to the total household income in euros (X). It is known that the specification of the Engel curve is a Working-Leser model,

$$Y = \beta_0 + \beta_1 \ln X + u \text{ with } \mathbb{E}(u|X) = 0.$$

a. How are the model coefficients affected when the household income is measured in thousands of euros? (100%)

b. How are the model coefficients when we take logarithms in base 10 instead of natural logarithms for X? (100%)

c. Suppose that the model is estimated with X in levels. That is, we estimate the model

$$Y = \delta_0 + \delta_1 X + v.$$

Provide an expression for the asymptotic bias of the OLS estimate of δ_1 as an estimate of β_1. (100%)

ANSWER:

a. The observed income is $X^* = X/1000$ and therefore $X = 1.000 \cdot X^*$. Then,

$$\begin{aligned} Y &= \beta_0 + \beta_1 \ln(1.000 \cdot X^*) + u \\ &= (\beta_0 + \beta_1 \ln(1.000)) + \beta_1 \ln X^* + u. \end{aligned}$$

Therefore, the slope of the model does not change, but the intercept is now $(\beta_0 + \beta_1 \ln(1.000))$.

b. $Z = \ln X \Rightarrow X = e^Z \Rightarrow \log_{10} X = Z \log_{10} e \Rightarrow Z = \log_{10} X / \log_{10} e$. Therefore,

$$Y = \beta_0 + \frac{\beta_1}{\log_{10} e} \log_{10} X + u.$$

The slope is now $\beta_1 / \log_{10} e$.

c. The OLS estimate of δ_1,

$$\begin{aligned} \hat{\delta}_1 &= \frac{\widehat{Cov}(X,Y)}{\widehat{Var}(X)} = \frac{\widehat{Cov}(X, \beta_0 + \beta_1 \ln X + u)}{\widehat{Var}(X)} \\ &= \beta_1 \frac{\widehat{Cov}(X, \ln X)}{\widehat{Var}(X)} + \frac{\widehat{Cov}(X,u)}{\widehat{Var}(X)}, \end{aligned}$$

replacing Y by the true model. Then the OLS estimate converges for large samples as

$$\hat{\delta}_1 \to_p \beta_1 \frac{Cov(X, \ln X)}{Var(X)} + \frac{Cov(X,u)}{Var(X)} = \beta_1 \frac{Cov(X, \ln X)}{Var(X)}$$

because $\mathbb{E}(u|X) = 0$ implies $Cov(X,u) = 0$. Then, the asymptotic bias of $\hat{\delta}_1$ is

$$\begin{aligned} \beta_1 \frac{Cov(X, \ln X)}{Var(X)} - \beta_1 &= \beta_1 \left(\frac{Cov(X, \ln X)}{Var(X)} - 1 \right) \\ &= \beta_1 \left(\frac{Cov(X, \ln X) - Var(X)}{Var(X)} \right) \\ &= \beta_1 \left(\frac{Cov(X, \ln X - X)}{Var(X)} \right). \end{aligned}$$

NAME: ______________________________ GROUP: __________

QUESTION 2 (20%): We want to estimate the price elasticity of the demand for milk. We have annual observations of quantities (Q) and prices (P). We also have data on the annual rainfall (R). We propose a supply and demand model where Q and P are in logarithms.

a. Show that the OLS estimate of the price elasticity in the demand curve is necessarily inconsistent. (100%)

b. Propose an instrumental variable to estimate the price elasticity of demand (20%) and explain how you would test that this instrumental variable is relevant using a t statistic (80%).

c. What restriction must the structural forms coefficients fulfill for the proposed instrumental variable to be relevant? (100%)

ANSWER:

a.

$$\begin{aligned} \text{Demand curve:} \quad & \ln Q = \beta_0 + \beta_1 \ln P + u_1 \\ \text{Supply curve:} \quad & \ln Q = \delta_0 + \delta_1 \ln P + \delta_2 R + u_2. \end{aligned}$$

Therefore,

$$\begin{aligned} 0 &= (\beta_0 - \delta_0) + (\beta_1 - \delta_1) \ln P - \delta_2 R + (u_1 - u_2) \\ \ln P &= \frac{\delta_0 - \beta_0}{\beta_1 - \delta_1} + \frac{\delta_2}{\beta_1 - \delta_1} R + \frac{u_1 - u_2}{\beta_1 - \delta_1}, \text{ assuming } \beta_1 \neq \delta_1. \end{aligned}$$

and $Cov(\ln P, u_1) = Var(u_1) / (\beta_1 - \delta_1) \neq 0$, so that the OLS estimate of the price elasticity in the demand curve,

$$\begin{aligned} \hat{\beta}_1 &= \frac{\widehat{Cov}(\ln P, \ln Q)}{\widehat{Var}(\ln P)} = \frac{\widehat{Cov}(\ln P, \beta_0 + \beta_1 \ln P + u_1)}{\widehat{Var}(\ln P)} \\ &= \beta_1 + \frac{\widehat{Cov}(\ln P, u_1)}{\widehat{Var}(\ln P)} \end{aligned}$$

is biased for β_1.

b. We estimate the reduced form for $\ln P$

$$\ln P = \pi_0 + \pi_1 R + v$$

and use the t statistic

$$t = \frac{\hat{\pi}_1}{SE(\hat{\pi}_1)}$$

to test

$$H_0 : \pi_1 = 0 \text{ vs } H_1 : \pi_1 \neq 0.$$

We reject at the $(1 - \alpha) 100\%$ significance level when $|t| > Z_{\alpha/2}$, with $\Pr(N(0,1) > Z_a) = a$.

c. δ_2 must be different from zero, because

$$\pi_1 = \frac{\delta_2}{\beta_1 - \delta_1}.$$

NAME: ______________________________ GROUP: __________

QUESTION 3 (20%): Consider a regression model with dependent variable Y and two explanatory variables, X_1 and X_2 (and the intercept).

a. Obtain and interpret the first-order conditions of the OLS estimates of the model parameters (20%). Then use these equations to derive an expression for the OLS estimate of the coefficient of X_1 (80%).

b. Suppose that $X_{1i} = \alpha_0 + \alpha_1 X_{2i}$, $i = 1, ..., n$, where α_0 and α_1 are constants different from zero. Which consequences this would have on the estimate in a.? (100%)

c. Suppose that the model is estimated without intercept. Which restrictions must be imposed on Y, X_1 and X_2 for consistency of the X_1 and X_2 coefficients? (100%)

ANSWER:

a. Class problem. The OLS estimate of β_1 is

$$\hat{\beta}_1 = \frac{\widehat{Cov}\left(Y, \hat{e}_1\right)}{\widehat{Var}\left(\hat{e}_1\right)},$$

where

$$\hat{e}_{1i} = X_{1i} - \hat{\alpha}_0 - \hat{\alpha}_1 X_{2i}$$

are the residuals from the regression

$$X_{1i} = \alpha_0 + \alpha_1 X_{2i} + e_{1i}$$

b. $\widehat{Corr}\left(X_1, X_2\right) = sign\left(\alpha_1\right)$. Therefore, there is perfect multicolinearity. In this case $\widehat{Var}\left(\hat{e}_1\right) = 0$, and the OLS estimate of β_1 can not be computed.

c. They are consistent when the intercept is actually equal to zero. We know that

$$\beta_0 = \mathbb{E}\left(Y\right) - \beta_1 \mathbb{E}\left(X_1\right) - \beta_2 \mathbb{E}\left(X_2\right).$$

Therefore, the OLS estimates are consistent when $\mathbb{E}\left(Y\right) = \mathbb{E}\left(X_1\right) = \mathbb{E}\left(X_2\right) = 0$.

QUESTION 4 WITH GRETL (40%): Use the file **CARD** from Wooldridge containing a sample of 3010 men. We are interested in estimating the returns to education for men in the United States, by regressing $\ln(wage)$ on $educ$, $exper$, $exper^2$, $black$, $SMSA$ and $south$, where $wage$ is the weekly wage in dollar cents, $educ$ is years of education, $exper$ is years of experience, $black = 1$ if the person is black, $SMSA = 1$ if the person lives in an Standard Metropolitan Statistical Area and $south = 1$ if he lives in a Southern state. You can assume that the variables $exper$, $exper^2$, $black$, $SMSA$ and $south$ are exogenous.

a. Argue why *educ* can be an endogenous explanatory variable in this regression (30%) and which are the consequences on the OLS estimates if so (20%). Explain what variables can be considered exogenous and what variables can be considered endogenous among the following: mother's education (*motheduc*), having lived near a university before starting work ($nearc4$), and the intelligence quotient (IQ). (50% mismo peso a cada variable)

b. Test that *motheduc* and $nearc4$ are relevant instruments, reporting the model used to perform the test (20%), the hypothesis to be tested on the parameters of interest (20%), the statistic to use in the test (20%) and the criterion to reject or not the hypothesis tested (20%). Are the instruments weak? (20%)

c. Can you test that *motheduc* and $nearc4$ are exogenous? (20%) If you answer is positive, perform the test: report the auxiliary model used (20%), the hypothesis to be tested on the parameters of interest (20%), the statistic to use in the test (20%) and the criterion to reject or not the hypothesis tested (20%).

d. Obtain the Two Stage Least Squares estimates of your model using *motheduc* and $nearc4$ as instruments (40%) and interpret the coefficient of *educ* (20%). Test that the returns to education are larger than 10% per additional year of education (40%).

ANSWER:

a. The obvious reason for *educ* endogeneity is the omitted ability story, which is a direct determinant of *wage*, but also of *educ*, leading to the correlation between *educ* and U.

   - Mother's education (*motheduc*): it could be exogenous if we think that mother's education can affect the education level of the children, but not the ability (because e.g. there is not a genetic link between the abilities of mother and children). However, it could be endogenous if we think that the ability of the individual, even if it is not a genetic characteristic, can be trained by parents involvement in education, which could be related with their education level.

   - Having lived near a university before starting work ($nearc4$) : it could be exogenous as in principle is not related to the ability of the individual (despite it could be correlated with the level of education of the individual as it makes cheaper to attend college).

   - Intelligence quotient (IQ) : endogenous because this is related (but not exactly equal) to ability.

   .

b. For checking relevance we have to run the 1st stage regression which regress the endogenous regressor in the structural equation in all the exogenous regressors and the instruments:

$$\begin{aligned} educ &= \pi_0 + \pi_1 exper + \pi_2 exper^2 + \pi_3 black + \pi_4 SMSA \\ &\quad + \pi_5 south + \pi_6 motheduc + \pi_7 nearc4 + V \end{aligned}$$

   and test the significance of the coefficients of the instruments

$$\begin{aligned} H_0 &: \pi_6 = \pi_7 = 0 \\ H_1 &: \pi_6 \neq 0 \text{ and/or } \pi_7 \neq 0 \end{aligned}$$

   through an F test. Using heteroskedasticity robust s.e.'s Gretl reports for this test

[Wald Test] F robust(2, 2649) = 97.2012, p-value 1.83314e-41,

so we reject the null hypothesis of no relevance in favour of the alternative of relevance. Since $F > 10$ we conclude that the instruments are not weak.

.

Under a conditional homoskedasticity assumption we could use the F test based on the comparison of the R^2 of the previous regression and the restricted one,

$$educ = \pi_0 + \pi_1 exper + \pi_2 exper^2 + \pi_3 black + \pi_4 SMSA + \pi_5 south + V$$

i.e. for $k = 7$ and $m = 2$,

$$F = \frac{n-k-1}{m}\frac{R^2_{unres} - R^2_{res}}{1 - R^2_{unres}} = \frac{2657-7-1}{2}\frac{0.4944 - 0.4509}{1 - 0.4944} = 113.96$$

and compare F with the 1% critical value from a $\chi^2_m/m = \chi^2_2/2$ distribution or checking that the $p - value$ $\approx 0 < 0.01$, so the null hypothesis is rejected and we can conclude that the instruments are relevant and since $F > 10$ they are not weak instruments.

.

c. Since we have an overidentified case, $2 = m > k = 1$, we can perform a test of instrument exogeneity using the J test. Using the 2SLS Gretl's output (which assumes homoskedasticity) for the structural equation

$$\ln(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + \beta_4 black + \beta_5 SMSA + \beta_6 south + U$$

we have:

.

Sargan Overidentification Test:

Null hypothesis: [All instruments are valid]

Statistic of Test: LM = 0.00671297

P-value = P(Chi-square(1) > 0.00671297) = 0.9347

.

Calculating directly the non-robust F test of significance of the instruments *motheduc* and *nearc*4,

$$\begin{aligned} H_0 &: \delta_{motheduc} = \delta_{nearc4} = 0 \\ H_0 &: \delta_{motheduc} \neq 0 \text{ and/or } \delta_{nearc4} \neq 0 \end{aligned}$$

in the auxiliary regression of the 2SLS residuals $\hat{U}$ on all exogenous variable

$$\hat{U} = \delta_0 + \delta_1 exper + \delta_2 exper^2 + \delta_3 black + \delta_4 SMSA + \delta_5 south + \delta_6 motheduc + \delta_7 nearc4 + e,$$

we have $F = 0.00334639$, so $J = mF = 2 \times 0.00334639 = 6.6928 \times 10^{-3}$, while the robust F test gives $F = 0.00348109$, and $J = 2 \times 0.00348109 = 6.9622 \times 10^{-3}$.

In any case, when compared to the critical value from a χ^2_1, 3.84, the null hypothesis of exogeneity is not rejected, and both instruments seem valid.

.

d. Gretl output is

.

Model 1: 2SLS, using observations 2–3006 ($n = 2657$)
Missing or incomplete observations are eliminated: 348
Dependent variable: lnwage
With instruments: educ
Instruments: const nearc4 motheduc black smsa exper expersq south
Robust standard errors against heteroskedasticity, HC1

	Coefficient	s.e	t statistic	p value
const	4.27057	0.232385	18.38	0.0000
educ	0.101267	0.0137082	7.387	0.0000
exper	0.0944650	0.00885729	10.67	0.0000
expersq	−0.00219596	0.000354182	−6.200	0.0000
black	−0.168645	0.0239724	−7.035	0.0000
smsa	0.146323	0.0177702	8.234	0.0000
south	−0.115911	0.0170531	−6.797	0.0000
	.	.	.	.

Mean dependent var.	6.270610	S.D dependent var.	0.444069
RSS	379.8867	S.E. of regression	0.378620
R^2	0.281480	R^2 adjusted	0.279853
$F(6, 2650)$	130.7680	P value (F)	2.7e–145

The coefficient of *educ*, $\hat{\beta}_{educ} = 0.101267$ it is interpreted as wages increasing on average a 10.12% when education increases in one year, everything else fixed. The hypotheses to be tested then are

$$\begin{aligned} H_0 &: \beta_{educ} = 0.10 \\ H_1 &: \beta_{educ} > 0.10 \end{aligned}$$

so the t statistic using the robust 2SLS standard errors is

$$t = \frac{\hat{\beta}_{educ} - 0.10}{SE\left(\hat{\beta}_{educ}\right)} = \frac{0.101267 - 0.10}{0.0137082} = 0.0924,$$

and we can not reject the null hypothesis at the usual significance levels, and the empirical evidence does not support that returns to education are larger than 10%.